

FINAL
02-07-2479
MP

Bandwidth Efficient Trellis-Coded Modulation
with Prescribed Decoding Delay - New Interpretations and Results

Marvin K. Simon, Scott Darden, Matthew Fong
Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109
California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125

Abstract

Motivated by previous work of Li and Rimoldi for obtaining bandwidth efficient TCM signals with finite decoding delay, we present an alternative representation for their encoder/signal mapper transmitter structure which merely consists of a single filter (with complex impulse response) having an input equal to the $(+1, -1)$ equivalent of the $(0,1)$ input data bits in their implementation. The filter impulse response is of duration $(\nu+1)T_b$ (ν is memory of the modulation, T_b is the bit time, and νT_b is the decoding delay) and can be constructed by designing its $\nu+1$ bit time partitions in terms of the waveform differences that characterize the finite decoding delay conditions found by Li and Rimoldi. The advantage of this simpler transmitter structure is that it readily allows computation of the modulation's power spectral density from which one can determine the conditions that must be imposed on the signal design to produce an equivalent low pass power spectral density. This in turn allows for a straightforward procedure for designing the optimum signals to produce maximum bandwidth efficiency as measured by fractional out-of-band power. Such optimum signal designs are determined for memory one and memory two modulations and presented as examples of the application of the general results.

Bandwidth Efficient Trellis-Coded Modulation with Prescribed Decoding Delay – New Interpretations and Results

Marvin K. Simon, Scott Darden, Matthew Fong

Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109
California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125

1. Introduction

In a paper presented at the 1997 International Symposium on Information Theory [1], Li and Rimoldi presented a particular transmitter structure (the combination of an encoder of memory v and a waveform mapper - see Fig. 1) for trellis-coded modulations (TCMs) that under certain constraints placed on the differences of the transmitted waveforms guaranteed decoding (using a conventional trellis decoder) with a finite (v bit duration) delay. Specifically, the encoder was simply a tapped delay line whose v taps together with the input bit were mapped into a set of $M = 2^{v+1}$ waveforms (signals) of one bit duration (T_b) in accordance with a binary coded decimal (BCD) relationship. That is, if $U_n \in 0,1$ denotes the n th input bit and $U_{n-1}, U_{n-2}, \dots, U_{n-v}$ the previous v bits (the state of the encoder), then the signal transmitted in the interval $nT_b \leq t \leq (n+1)T_b$ would be $s_i(t)$ where the index i is defined in terms of these bits by $i = U_n \times 2^v + U_{n-1} \times 2^{v-1} + \dots + U_{n-v+1} \times 2^1 + U_{n-v} \times 2^0$. It was also shown in [1] that, in addition to the constraints placed on the waveform differences, it was possible to further constrain the signals so as to maximize the value of the minimum squared Euclidean distance taken over all pairs of error event paths, namely, $d_{\min}^2 = 2$. Such a maximum value of d_{\min}^2 , which corresponds to a number of binary modulations such as binary phase-shift-keying (BPSK) and the more bandwidth efficient minimum-shift-keying (MSK), indicates that the receiver is providing optimum reception from a power conservation standpoint. Finally, in the presence of all of the above constraints, Li and Rimoldi [1] showed that it is possible to further optimize the system by selecting a set of waveforms that minimize the bandwidth-bit time product BT_b .

In this paper, we investigate an alternative (simpler) representation of the transmitter configuration suggested in [1] which consists of nothing more than a single filter (with complex impulse response) whose input is the ± 1 equivalent of the input data bits, namely, $\bar{U}_n = 1 - 2U_n$ for all n . This representation comes about by viewing the transmitted signal as a random pulse train with a pulse shape that extends beyond a single bit interval, i.e., one that contributes intersymbol interference (ISI) to its neighbors. As we shall see, such a pulse shape of duration

$(v+1)T_b$ can be constructed by designing its $v+1$ partitions of duration T_b sec in terms of the waveform differences that are outputted from Li and Rimoldi's transmitter. Such an ISI-based transmitter representation has the advantage that the power spectral density (PSD) and hence the bandwidth are readily evaluated using known results for uncoded random binary complex pulse trains. It also allows applying the insight provided in Forney's classic paper [2] on the Viterbi algorithm, in particular the discussion regarding the use of this algorithm to combat ISI.

One of the requirements placed on the set of possible transmitted waveforms $s_i(t), i = 0, 1, \dots, M$ in [1] is that they all have equal energy.¹ The impact of relaxing the equal energy restriction on the power efficiency of the modulation scheme in its ability to achieve the largest value of d_{\min}^2 has also been investigated but is omitted here in the interest of brevity. Suffice it to say that with an additional set of constraints (now on the differences of the *energies* of the signals) that must be satisfied to achieve the same finite decoding delay, the optimum sequence receiver results in a signal design with a maximum value of d_{\min}^2 less than two. Allowing the signals to have unequal energy, however, suggests the possibility of additional flexibility in the design of these signals in order to achieve the best bandwidth efficiency, i.e., the reduction in d_{\min}^2 caused by the unequal energy requirement can possibly trade off against an additional reduction in signal bandwidth. Additional consideration of this notion warrants investigation.

2. ISI-Based Transmitter Implementation

The decomposition of a memory modulation into a cascade of an encoder and a memoryless modulator was first applied to continuous phase modulation (CPM) by Rimoldi [3]. In particular, for MSK, a special case of CPM corresponding to a rectangular frequency pulse of duration T_b sec (full response) and frequency modulation index (two-sided frequency deviation normalized by the bit rate) $h = 0.5$, the memory is $v = 1$ and a transmitter analogous to Fig. 1 was obtained as in Fig. 2. Comparing Figs. 1 and 2 we note that in the latter the state is represented by the *differentially encoded* version of the current input bit $V_n = U_n \oplus V_{n-1}$ whereas in the former it would be just the previous input bit U_{n-1} itself. Furthermore, because of the differential encoding associated with the state in Fig. 2, a differential decoder

¹Note that the assumption of equal energy does not imply constant envelope as was the case for the continuous phase modulations (CPMs) studied in [3] which served as the motivation for the work leading up to the results in [1]. Nevertheless, the envelope fluctuation of the resulting signal designs will be small when compared with Nyquist designs of comparable bandwidth efficiencies.

would be required in the receiver following the trellis decoder which results in a small loss in bit error probability (BEP) performance. It is well known [4, Chap. 10] that precoding true MSK with a differential decoder at the transmitter results in a modulation that is equivalent (spectral and power efficiently) to MSK but without the need for differential decoding at the receiver. It is such precoded MSK that is implemented by the simpler configuration of Fig. 1. In what follows, when referring to MSK in the context of Fig. 1 or its equivalents, we shall assume that precoded MSK is what is implied.

Consider an uncoded random binary (± 1) sequence $\{\bar{U}_n\}$ which generates a random pulse train

$$s'(t) = \sum_{n=-\infty}^{\infty} \bar{U}_n p(t - nT_b) \quad (1)$$

where $p(t) \triangleq p_R(t) + jp_I(t)$ is a complex pulse shape defined on the interval $0 \leq t \leq (v+1)T_b$. Consider partitioning $p(t)$ into $v+1$ adjoint pieces corresponding to its one-bit interval sections. That is we define the set of T_b -sec duration waveforms

$$p_k(t) \triangleq p_{Rk}(t) + jp_{Ik}(t) = \begin{cases} p(t + kT_b), & 0 \leq t \leq T_b, \\ 0, & \text{otherwise} \end{cases}, \quad k = 0, 1, 2, \dots, v \quad (2)$$

From (1), in any T_b -sec interval, e.g., the n th, the signal $s'(t)$ will be described by one of $M = 2^{v+1}$ complex waveforms, i.e., $s'_k(t - nT_b)$, $k = 0, 1, 2, \dots, 2^{v+1} - 1$, which are expressed in terms of $p(t)$ and the data sequence $\{\bar{U}_n\}$ by

$$s'_k(t - nT_b) = \bar{U}_n p_0(t - nT_b) + \bar{U}_{n-1} p_1(t - nT_b) + \dots + \bar{U}_{n-v} p_v(t - nT_b), \quad k = 0, 1, 2, \dots, 2^{v+1} - 1 \quad (3)$$

where the index k is the equivalent (0,1) bit sequence $\{U_n, U_{n-1}, \dots, U_{n-v}\}$ expressed in BCD form. As an example, the set of waveforms for memory $v = 2$ is given below:

$$\begin{aligned} s'_0(t - nT_b) &= p_0(t - nT_b) + p_1(t - nT_b) + p_2(t - nT_b), & s'_1(t - nT_b) &= p_0(t - nT_b) + p_1(t - nT_b) - p_2(t - nT_b) \\ s'_2(t - nT_b) &= p_0(t - nT_b) - p_1(t - nT_b) + p_2(t - nT_b), & s'_3(t - nT_b) &= p_0(t - nT_b) - p_1(t - nT_b) - p_2(t - nT_b) \\ s'_4(t - nT_b) &= -p_0(t - nT_b) + p_1(t - nT_b) + p_2(t - nT_b), & s'_5(t - nT_b) &= -p_0(t - nT_b) + p_1(t - nT_b) - p_2(t - nT_b) \\ s'_6(t - nT_b) &= -p_0(t - nT_b) - p_1(t - nT_b) + p_2(t - nT_b), & s'_7(t - nT_b) &= -p_0(t - nT_b) - p_1(t - nT_b) - p_2(t - nT_b) \end{aligned} \quad (4)$$

We note from (4) that because of the BCD construction, the following properties hold for the signal differences:

$$s'_0(t) - s'_1(t) = s'_2(t) - s'_3(t) = s'_4(t) - s'_5(t) = s'_6(t) - s'_7(t) = 2p_2(t) \quad (5a)$$

$$s'_0(t) - s'_2(t) = s'_4(t) - s'_6(t) = 2p_1(t) \quad (5b)$$

Also, an equivalent (at least in so far as the first equality is concerned) condition to (5b) is

$$s'_0(t) - s'_4(t) = s'_2(t) - s'_6(t) = 2p_0(t) \quad (5c)$$

In the more generic case for arbitrary v , the conditions corresponding to (5a) and (5b) would be summarized as:

$$s'_0(t) - s'_{2^m}(t) = s'_{2^{m+1}l}(t) - s'_{2^{m+1}l+2^m}(t) = 2p_{v-m}(t), \quad m = 0, 1, 2, \dots, v-1, \quad l = 1, 2, \dots, 2^{v-m} - 1 \quad (6)$$

and in addition the generalization of (5c) becomes

$$s'_0(t) - s'_{2^v}(t) = s'_{2^{v-1}}(t) - s'_{2^v+2^{v-1}}(t) = 2p_0(t) \quad (7)$$

Associating the 2^{v+1} signals $\{s'_k(t)\}$ expressed as in (3) with the assumed equal energy $\{s_k(t)\}$ derived from the implementation in Fig. 1, we see that the conditions on the signal differences of $s'_l(t)$ given in (6) are precisely those of Theorem I in [1] which guarantees a finite decoding delay of v bits using an optimum trellis-coded receiver.² Therefore, since $p(t)$ is entirely specified by its adjoint T_b -sec sections $p_i(t), i = 0, 1, \dots, v$, it would appear that the transmitter of Fig. 1 can be equivalently implemented (see Fig. 3) by passing the input ± 1 data sequence $\{\bar{U}_n\}$ (modeled as a random impulse train) through a filter with complex impulse response

$$p(t) = \sum_{i=0}^v p_i(t - iT_b), \quad p_i(t) = \frac{1}{2} [s'_0(t) - s'_{2^{v-i}}(t)] \quad (8)$$

or equivalently, the real and imaginary parts of the baseband signal (to be modulated onto quadrature carriers for transmission over the channel) can be obtained by passing the common input ± 1 data sequence $\{\bar{U}_n\}$ through a pair of filters with respective impulse responses

²It has also been noted by Li and Rimoldi that these conditions guarantee that the Euclidean distance between any pair of paths in the trellis decoder that diverge at time n and remerge at time $n+v+1$ is the same. Furthermore, the number of correlators (matched filters) needed to implement the optimum (MLSE) receiver will now vary *linearly* with memory, i.e., $v+1$, as opposed to *exponentially* with memory, i.e., 2^{v+1} , which is the case when no constraints are imposed on the decoding delay.

$$p_{Ri}(t) = \frac{1}{2}[s'_{R0}(t) - s'_{R2^{v-i}}(t)], \quad p_{Li}(t) = \frac{1}{2}[s'_{i0}(t) - s'_{i2^{v-i}}(t)] \quad (9)$$

Unfortunately, the implementation in Fig. 3 is not always equivalent to that in Fig. 1, but as we shall see momentarily, for the case of most practical interest, i.e., a signal set $\{s_k(t)\}$ with maximum minimum Euclidean distance between its members, the equivalence between the two implementations is guaranteed, i.e., $\{s'_k(t)\}$ and $\{s_k(t)\}$ are identical. Before showing this, we note that even though $p_R(t)$ and $p_L(t)$ are constructed from the real and imaginary components of a set of equal energy complex signals $\{s'_k(t), k = 0, 1, 2, \dots, 2^{v+1} - 1\}$, they themselves do not necessarily have equal energy. We shall see that this is true even for the simple case of MSK.

Note that because of the symmetry of the BCD mapping, the signals in the memory 2 example of (4) also satisfy the conditions

$$s'_0(t) = -s'_7(t), s'_1(t) = -s'_6(t), s'_2(t) = -s'_5(t), s'_3(t) = -s'_4(t) \quad (10)$$

which in the case of arbitrary memory v would become

$$s'_m(t) = -s'_{2^{v+1}-1-m}(t), m = 0, 1, \dots, 2^v - 1 \quad (11)$$

The conditions of (11) which correspond to an antipodal signaling set are precisely those given in [1] that achieve the maximum value of minimum squared Euclidean distance, namely, $d_{\min}^2 = 2$. Thus, the implementation of Fig. 3 not only achieves finite decoding delay but also automatically achieves the optimum performance from the standpoint of power efficiency. This result should not be surprising in view of the findings in [2] which indicate that a maximum-likelihood (optimum) sequence estimator (MLSE) form of receiver such as the trellis decoder can completely remove the ISI and thereby achieve the performance of a zero ISI (full response) system. However, since the implementation of Fig. 1 can produce a set of signals $\{s_k(t)\}$ that satisfy the difference properties needed for finite decoding delay without requiring them to have maximum minimum Euclidean distance, then the two implementations will be equivalent, i.e., $\{s_k(t)\} = \{s'_k(t)\}$ only when this additional requirement is imposed. A formal proof of this equivalence is discussed in the longer version of the Li and Rimoldi paper in [1]. In what follows, we consider only the practically important case of antipodal signal sets and as such drop the prime notation on the signals derived from $p(t)$.

What remains is to consider the bandwidth efficiency of signals designed according to the constraints of (6), (7), and (11). This is where the ISI-based representation of Fig. 3 helps considerably since the evaluation of the PSD of the

transmitted signal can be trivially accomplished using well-known relations [4] for random pulse trains. This is considered in the next section.

3. Evaluation of the Power Spectral Density

In this section we compute the PSD of a random complex pulse train, e.g., that in (1), modulated onto quadrature carriers. That is, if the transmitted bandpass signal is given by³

$$\tilde{s}(t) = \text{Re}\{s(t)e^{j2\pi f_m t}\} = \left(\sum_{n=-\infty}^{\infty} \bar{U}_n p_R(t - nT_b) \right) \cos 2\pi f_m t - \left(\sum_{n=-\infty}^{\infty} \bar{U}_n p_I(t - nT_b) \right) \sin 2\pi f_m t \quad (12)$$

then it is straightforward to show using an extension of the methods in [4, Chap. 2] that the PSD of $\tilde{s}(t)$ is given by

$$\begin{aligned} S(f) &= \frac{1}{4T_b} |P_R(f - f_m) + jP_I(f - f_m)|^2 + \frac{1}{4T_b} |P_R(f + f_m) - jP_I(f + f_m)|^2 \\ &= S_u(f) + S_l(f) \end{aligned} \quad (13)$$

where

$$P_R(f) \triangleq \mathcal{F}\{p_R(t)\}, \quad P_I(f) \triangleq \mathcal{F}\{p_I(t)\} \quad (14)$$

are the Fourier transforms of the real and imaginary pulse shapes which, in general, are complex functions of f and the u and l subscripts denote upper and lower sideband, respectively. Note that the signal in (12) differs from the usual quadrature phase-shift-keying (QPSK)-type of signal in that here the same data sequence is passed through both the inphase (I) and quadrature (Q) filters whereas for QPSK the two sequences passing through these filters would be different and independent of one another. As such, the PSD in (13) cannot, in general, be written in the form [4, Chap. 2, Eq. (2.131)].

$$S(f) = \frac{1}{4} G(f - f_c) + \frac{1}{4} G(f + f_c) \quad (15)$$

where $G(f)$ is the equivalent baseband (symmetrical around $f = 0$) PSD and is a real function of f and f_c is some arbitrary carrier frequency.⁴

³We use the notation " f_m " for the actual modulating frequency of the quadrature carriers to distinguish it from the carrier frequency around which the PSD is symmetric which will be denoted by " f_c ". More about this shortly.

⁴What is meant by an "equivalent baseband PSD" is a PSD around zero frequency that is *identical* to the upper or lower sideband of the bandpass PSD frequency-shifted to the origin. While it is always possible to obtain a symmetric PSD around the origin by demodulating the bandpass signal with a

To illustrate the above point consider the specific case of MSK ($\nu = 1$) for which the four complex signals are given by⁵

$$s_0(t) = 0 + j1, \quad s_1(t) = \sin \frac{\pi t}{T_b} - j \cos \frac{\pi t}{T_b} = s_0^*(t) e^{j \frac{\pi}{T_b} t} \quad (16)$$

$$s_2(t) = -s_1(t), \quad s_3(t) = -s_0(t)$$

In terms of the ISI-based representation, we obtain from (8) that

$$p_0(t) = \frac{1}{2} \sin \frac{\pi t}{T_b} + j \frac{1}{2} \left[1 - \cos \frac{\pi t}{T_b} \right] \quad (17)$$

$$p_1(t) = -\frac{1}{2} \sin \frac{\pi t}{T_b} + j \frac{1}{2} \left[1 + \cos \frac{\pi t}{T_b} \right]$$

Thus, using (17) to define the complex pulse shape of (8), we obtain

$$p(t) = \frac{1}{2} \sin \frac{\pi t}{T_b} + j \frac{1}{2} \left[1 - \cos \frac{\pi t}{T_b} \right], \quad 0 \leq t \leq 2T_b \quad (18)$$

That is, an appropriate implementation for MSK which guarantees a decoding delay of one bit is that of Fig. 3 with I and Q filters having impulse responses

$$p_R(t) = \frac{1}{2} \sin \frac{\pi t}{T_b}, \quad 0 \leq t \leq 2T_b \quad (19)$$

$$p_I(t) = \frac{1}{2} \left[1 - \cos \frac{\pi t}{T_b} \right], \quad 0 \leq t \leq 2T_b$$

Taking the Fourier transforms of $p_R(t)$ and $p_I(t)$ of (8) and using these in (13) we arrive at the following result for the bandpass PSD:

$$S(f) = \frac{T_b}{4} \frac{\sin^2 2\pi(f - f_m)T_b}{\pi^2} \left[\frac{1}{1 - 2(f - f_m)T_b} + \frac{1}{2(f - f_m)T_b} \right]^2$$

$$+ \frac{T_b}{4} \frac{\sin^2 2\pi(f + f_m)T_b}{\pi^2} \left[\frac{1}{1 + 2(f + f_m)T_b} - \frac{1}{2(f + f_m)T_b} \right]^2 = S_u(f) + S_l(f) \quad (20)$$

carrier at some frequency f_c (not necessarily equal to the modulating frequency f_m), the resulting PSD will, in general, be a combination (sum) of the aliased upper and lower sidebands and depending on the relative spacing between f_c and f_m may or may not appear as a simple frequency translation of either of these sidebands.

⁵Note that for the Rimoldi decomposition of MSK illustrated in Fig. 2, the signals satisfy the condition $s_0(t) - s_1(t) = -(s_2(t) - s_3(t))$ rather than $s_0(t) - s_1(t) = s_2(t) - s_3(t)$ as in (5a).

Note that while $S(f)$ is an even function of f (as it should be for a real signal), its upper and lower sidebands $S_u(f)$ and $S_l(f)$ are not symmetric around f_m and $-f_m$, respectively. However, there does exist a frequency, $f_c \neq f_m$, around which the upper sideband (and similarly for the lower sideband) is symmetric. To understand why this is so, we remind the reader that according to Rimoldi's decomposition [3], the modulation frequency chosen for the quadrature carriers should be shifted from the carrier frequency f_c around which the bandpass spectrum is to be symmetric by an amount equal to $1/4T_b$, i.e., $f_m = f_c - 1/4T_b$. The reason for this stems from the fact that the specification of the signals as in (16) results in a tilted trellis where the phase tilt is equal to $\pi/2$ radians. (Note that a frequency shift of $\Delta f = 1/4T_b$ is equal to a phase shift $2\pi\Delta f T = \pi/2$). To demonstrate that this is indeed the case, we evaluate the PSD of MSK using (20) with the shifted value of modulating frequency $f_m = f_c - 1/4T_b$. When this is done the result in (15) is obtained with

$$G(f) = \frac{16T_b}{\pi^2} \frac{\cos^2 2\pi f T_b}{(1 - 16f^2 T_b^2)^2} \quad (21)$$

which corresponds (except for a normalization factor) to the well-know PSD of MSK [4, Chap. 2, Eq. (2.148)].

The question now that comes about is: For arbitrary memory ν and a baseband signal design satisfying (6), (7) and (11), is it possible to find a modulating frequency f_m that will produce a symmetric bandpass PSD around some other carrier frequency f_c ? If not then one cannot find an equivalent baseband PSD and hence the bandwidth (whatever measure is used) of the signal must be determined from the RF waveform.

3.1 The Memory One Case

To shed some light on the answer to the above question, we consider the simplest case of unit memory where the complex pulse shape of (8) is simply given by

$$\begin{aligned} p(t) &= \frac{1}{2} [s_0(t) - s_2(t) + s_0(t - T_b) - s_1(t - T_b)] \\ &= \frac{1}{2} [s_0(t) + s_0(t - T_b) + s_1(t) + s_2(t - T_b)], \quad 0 \leq t \leq 2T_b \end{aligned} \quad (22)$$

where, in accordance with (11), we have used the fact that $s_1(t) = -s_2(t)$ in order to achieve $d_{\min}^2 = 2$. The Fourier transform of $p(t)$ in (22) is given by

$$P(f) = \frac{1}{2} \left[\int_0^{T_b} s_0(t) (1 + e^{-j2\pi f T_b}) e^{-j2\pi f t} dt + \int_0^{T_b} s_1(t) e^{-j2\pi f t} dt + e^{-j2\pi f T_b} \int_0^{T_b} s_2(t) e^{-j2\pi f t} dt \right] \quad (23)$$

Since from (13) the upper spectral sideband is $S_u(f) = \frac{1}{4T_b} |P(f - f_m)|^2$, then in order for this to be symmetric around f_c , we must have

$$|P(f_c + f - f_m)|^2 = |P(f_c - f - f_m)|^2 \quad (24)$$

or letting $f_s \triangleq f_c - f_m$ denote the separation between the actual modulation frequency and the bandpass frequency around which symmetry is desired, $s_0(t)$ and $s_1(t)$ must be chosen to satisfy

$$|P(f_s + f)|^2 = |P(f_s - f)|^2 \quad (25a)$$

or equivalently

$$|P(f_s + f)|^2 = |P^*(f_s - f)|^2 \quad (25b)$$

for some f_s . In terms of (23), the spectral equality in (25b) requires that we have

$$\begin{aligned} & \left| \int_0^{T_b} (s_0(t) e^{-j2\pi f_s t}) e^{-j2\pi f t} dt + e^{-j2\pi(f_s + f)T_b} \int_0^{T_b} (s_0(t) e^{-j2\pi f_s t}) e^{-j2\pi f t} dt \right. \\ & \left. + \int_0^{T_b} (s_1(t) e^{-j2\pi f_s t}) e^{-j2\pi f t} dt + e^{-j2\pi(f_s + f)T_b} \int_0^{T_b} (s_2(t) e^{-j2\pi f_s t}) e^{-j2\pi f t} dt \right|^2 \\ & = \left| \int_0^{T_b} (s_0^*(t) e^{j2\pi f_s t}) e^{-j2\pi f t} dt + e^{j2\pi(f_s - f)T_b} \int_0^{T_b} (s_0^*(t) e^{j2\pi f_s t}) e^{-j2\pi f t} dt \right. \\ & \left. + \int_0^{T_b} (s_1^*(t) e^{j2\pi f_s t}) e^{-j2\pi f t} dt + e^{j2\pi(f_s - f)T_b} \int_0^{T_b} (s_2^*(t) e^{j2\pi f_s t}) e^{-j2\pi f t} dt \right|^2 \end{aligned} \quad (26)$$

Sufficient conditions on the signals $\{s_i(t)\}$ for (26) to be satisfied are

$$s_1(t) = s_0^*(t) e^{j4\pi f_s t}, \quad s_2(t) = e^{j4\pi f_s T_b} s_0^*(t) e^{j4\pi f_s t} \quad (27)$$

However, since in arriving at (26) we have already assumed that $s_1(t) = -s_2(t)$, then (27) further requires that $f_s = 1/4T_b$ from which we obtain the complete signal set

$$s_1(t) = s_0^*(t) e^{j\pi t/T_b}, \quad s_2(t) = -s_0^*(t) e^{j\pi t/T_b}, \quad s_3(t) = -s_0(t) \quad (28)$$

Note that for memory one it is only necessary to specify $s_0(t)$ in order to arrive at the complete signal set. Also, the signal set of (28) satisfies the finite decoding delay condition of [1], namely, $s_0(t) - s_1(t) = s_2(t) - s_3(t)$.

The equivalent lowpass PSD is obtained by first using $s_1(t) = -s_2(t)$ in (23) resulting in

$$P(f) = \frac{1}{2} [S_0(f) + S_1(f) + e^{-j2\pi f T_b} (S_0(f) - S_1(f))] \quad (29)$$

from which one immediately gets

$$\frac{1}{T_b} |P(f)|^2 = \frac{1}{2T_b} [|S_0(f)|^2 + |S_1(f)|^2 + \text{Re}\{(S_0^*(f) + S_1^*(f))(S_0(f) - S_1(f))e^{-j2\pi f T_b}\}] \quad (30)$$

In (29) and (30), $S_i(f)$ denotes the Fourier transform of $s_i(t)$. Using the first symmetry condition of (28) in (30) gives the desired equivalent lowpass PSD, namely,

$$\begin{aligned} \frac{1}{T_b} |P(f + \frac{1}{4T_b})|^2 &= |S_0(f + \frac{1}{4T_b})|^2 [1 - \sin 2\pi f T_b] + |S_0(-f + \frac{1}{4T_b})|^2 [1 + \sin 2\pi f T_b] \\ &\quad + 2[\text{Re}\{S_0(f + \frac{1}{4T_b})\} \text{Im}\{S_0(-f + \frac{1}{4T_b})\} + \text{Re}\{S_0(-f + \frac{1}{4T_b})\} \text{Im}\{S_0(f + \frac{1}{4T_b})\}] \cos 2\pi f T_b \end{aligned} \quad (31)$$

which is clearly an even function of frequency.

Although (28) is satisfied by the MSK signals of (16) as should be the case, this condition applies in a more general context since it does not explicitly specify $s_0(t)$ but rather only the *relation between* $s_0(t)$ and $s_1(t)$. This should not be surprising since it has been shown in the past that there exists an entire class of MSK-type signals (referred to in [5] as generalized MSK) which happen to also be constant envelope (in addition to being equal energy) and achieve $d_{\min}^2 = 2$ as well as a decoding delay of one bit interval. In particular, the class of binary full response CPM signals with modulation index $h = 1/2$ and equivalent phase pulse $f(t)$ which satisfies the conditions $f(0) = 0, f(T_b) = 1/2$ is appropriate an example of which is Amoroso's *sinusoidal frequency-shift-keying (SFSK)* [6] for which

$$f(t) = \frac{t}{2T_b} \left(1 - \frac{\sin 2\pi t / T_b}{2\pi t / T_b} \right), \quad 0 \leq t \leq T_b \quad (32)$$

3.2 The Memory Two Case

For memory two, the pulse shape is given by

$$\begin{aligned} p(t) &= \frac{1}{2} [s_0(t) - s_4(t) + s_0(t - T_b) - s_2(t - T_b) + s_0(t - 2T_b) - s_1(t - 2T_b)] \\ &= \frac{1}{2} [s_0(t) + s_0(t - T_b) + s_0(t - 2T_b) + s_3(t) - s_2(t - T_b) - s_1(t - 2T_b)], \quad 0 \leq t \leq 3T_b \end{aligned}$$

(33)

with Fourier transform

$$P(f) = \frac{1}{2} \left[\left(1 + e^{-j2\pi f T_b} + e^{-j4\pi f T_b} \right) \int_0^{T_b} s_0(t) e^{-j2\pi f t} dt + \int_0^{T_b} s_3(t) e^{-j2\pi f t} dt \right. \\ \left. - e^{-j2\pi f T_b} \int_0^{T_b} s_2(t) e^{-j2\pi f t} dt - e^{-j4\pi f T_b} \int_0^{T_b} s_1(t) e^{-j2\pi f t} dt \right] \quad (34)$$

Applying (34) to (25b) and letting $s_3(t) = s_2(t) - s_0(t) + s_1(t)$ in accordance with (5a) we obtain the bandpass spectral symmetry condition

$$\left| e^{-j2\pi(f_s+f)T_b} \int_0^{T_b} (s_0(t) e^{-j2\pi f_s t}) e^{-j2\pi f t} dt + e^{-j4\pi(f_s+f)T_b} \int_0^{T_b} (s_0(t) e^{-j2\pi f_s t}) e^{-j2\pi f t} dt \right. \\ \left. + \int_0^{T_b} (s_2(t) e^{-j2\pi f_s t}) e^{-j2\pi f t} dt - e^{-j2\pi(f_s+f)T_b} \int_0^{T_b} (s_2(t) e^{-j2\pi f_s t}) e^{-j2\pi f t} dt \right. \\ \left. + \int_0^{T_b} (s_1(t) e^{-j2\pi f_s t}) e^{-j2\pi f t} dt - e^{-j4\pi(f_s+f)T_b} \int_0^{T_b} (s_1(t) e^{-j2\pi f_s t}) e^{-j2\pi f t} dt \right|^2 \\ = \left| e^{j2\pi(f_s-f)T_b} \int_0^{T_b} (s_0^*(t) e^{j2\pi f_s t}) e^{-j2\pi f t} dt + e^{j4\pi(f_s-f)T_b} \int_0^{T_b} (s_0^*(t) e^{j2\pi f_s t}) e^{-j2\pi f t} dt \right. \\ \left. + \int_0^{T_b} (s_2^*(t) e^{j2\pi f_s t}) e^{-j2\pi f t} dt - e^{j2\pi(f_s-f)T_b} \int_0^{T_b} (s_2^*(t) e^{j2\pi f_s t}) e^{-j2\pi f t} dt \right. \\ \left. + \int_0^{T_b} (s_1^*(t) e^{j2\pi f_s t}) e^{-j2\pi f t} dt - e^{j4\pi(f_s-f)T_b} \int_0^{T_b} (s_1^*(t) e^{j2\pi f_s t}) e^{-j2\pi f t} dt \right|^2 \quad (35)$$

Analogous with (27), satisfying (35) implies the set of conditions

$$s_1(t) + s_2(t) = (s_1^*(t) + s_2^*(t)) e^{j4\pi f_s t} \quad (36a)$$

$$s_0(t) - s_2(t) = e^{j4\pi f_s T_b} (s_0^*(t) - s_2^*(t)) e^{j4\pi f_s t} \quad (36b)$$

$$s_0(t) - s_1(t) = e^{j8\pi f_s T_b} (s_0^*(t) - s_1^*(t)) e^{j4\pi f_s t} \quad (36c)$$

Again letting $f_s = 1/4T_b$ and summing (36a), (36b) and (36c) gives

$$s_1(t) + s_2(t) = (s_1^*(t) + s_2^*(t)) e^{j\pi/T_b} \quad (37a)$$

$$s_0(t) = s_2^*(t) e^{j\pi/T_b} \quad (\text{or equivalently } s_2(t) = s_0^*(t) e^{j\pi/T_b}) \quad (37b)$$

$$s_0(t) - s_1(t) = (s_0^*(t) - s_1^*(t)) e^{j\pi/T_b} \quad (37c)$$

Actually, (37c) is not an independent condition since it can be derived from (37a) and (37b). Thus, (37a) and (37b) are sufficient to determine the signal design.

Following along the lines of (29) and (30) the equivalent PSD of the memory two modulation may be found. In particular, the Fourier transform of the equivalent pulse shape in (8) is given as

$$P(f) = \frac{1}{2} \left[S_0(f) + S_3(f) + e^{-j2\pi T_b} (S_0(f) - S_2(f)) + e^{-j4\pi T_b} (S_0(f) - S_1(f)) \right] \quad (38)$$

Using the additional relation $S_3(f) = S_1(f) + S_2(f) - S_0(f)$ to achieve finite decoding delay, one immediately gets the desired equivalent lowpass PSD as

$$\begin{aligned} \frac{1}{T_b} \left| P\left(f + \frac{1}{4T_b}\right) \right|^2 &= \frac{1}{4T_b} \left[\left| S_1\left(f + \frac{1}{4T_b}\right) + S_2\left(f + \frac{1}{4T_b}\right) \right|^2 + \left| S_0\left(f + \frac{1}{4T_b}\right) - S_2\left(f + \frac{1}{4T_b}\right) \right|^2 + \left| S_0\left(f + \frac{1}{4T_b}\right) - S_1\left(f + \frac{1}{4T_b}\right) \right|^2 \right. \\ &\quad + 2 \operatorname{Re} \left\{ \left(S_1^*\left(f + \frac{1}{4T_b}\right) + S_2^*\left(f + \frac{1}{4T_b}\right) \right) \left(S_0\left(f + \frac{1}{4T_b}\right) - S_2\left(f + \frac{1}{4T_b}\right) \right) e^{-j2\pi\left(f + \frac{1}{4T_b}\right)T_b} \right\} \\ &\quad + 2 \operatorname{Re} \left\{ \left(S_0^*\left(f + \frac{1}{4T_b}\right) - S_2^*\left(f + \frac{1}{4T_b}\right) \right) \left(S_0\left(f + \frac{1}{4T_b}\right) - S_1\left(f + \frac{1}{4T_b}\right) \right) e^{-2\pi\left(f + \frac{1}{4T_b}\right)T_b} \right\} \\ &\quad \left. + 2 \operatorname{Re} \left\{ \left(S_1^*\left(f + \frac{1}{4T_b}\right) + S_2^*\left(f + \frac{1}{4T_b}\right) \right) \left(S_0\left(f + \frac{1}{4T_b}\right) - S_1\left(f + \frac{1}{4T_b}\right) \right) e^{-4\pi\left(f + \frac{1}{4T_b}\right)T_b} \right\} \right] \end{aligned} \quad (39)$$

which using (37) can be shown to be an even function of frequency as is necessary.

4. Optimizing the Bandwidth Efficiency

Having now obtained expressions for the equivalent baseband PSD, it is now straightforward to use these to determine the sets of signals that satisfy all of the previous constraints and in addition maximize the power within a given bandwidth B . In mathematical terms, we search for the set of signals that for a given value of B maximizes the fractional in-band power

$$\eta = \frac{\int_{-B/2}^{B/2} G(f) df}{\int_{-\infty}^{\infty} G(f) df}, \quad G(f) \triangleq \frac{1}{T_b} \left| P\left(f + \frac{1}{4T_b}\right) \right|^2 \quad (40)$$

subject to the unit power constraint

$$\frac{1}{T_b} \int_0^{T_b} |s_i(t)|^2 dt = \frac{1}{T_b} \int_{-\infty}^{\infty} |S_i(f)|^2 df = 1, \quad i = 0, 1, 2, \dots, M-1 \quad (41)$$

4.1 Memory One Case

For the case of $v=1$ we observed that the entire signal set may be determined from the single complex signal $s_0(t)$. Thus, the optimization of bandwidth efficiency corresponds to substituting the PSD of (31) (which is entirely specified in terms of the Fourier transform of $s_0(t)$) into (40) and then maximizing η subject to (41). Such a procedure would result in an optimum $S_0(f)$ from whose inverse Fourier

transform one could determine the optimum signal set. Since $S_0(f)$ exists, in general, over the entire doubly infinite frequency axis, it is perhaps simpler to approach the optimization in the time domain since $s_0(t)$ is indeed time-limited to the interval $0 \leq t \leq T_b$. To do this we need to first rewrite the PSD of (41) in terms of $s_0(t)$ rather than $S_0(f)$ and then perform the integrations on f required in (40). After considerable manipulation, and for simplicity of notation normalizing $T_b = 1$ (i.e., $BT_b = B$), it can be shown that

$$\begin{aligned} \int_{-B/2}^{B/2} G(f)df &= B \int_0^1 \int_0^1 s_0(t) s_0^*(\tau) e^{-j\frac{\pi}{2}(t-\tau)} \left[\text{sinc } \pi B(t-\tau) - j \frac{1}{2} \text{sinc } \pi B(t-\tau+1) \right. \\ &\quad \left. + j \frac{1}{2} \text{sinc } \pi B(t-\tau-1) \right] dt d\tau \\ &\quad + \frac{1}{2} B \text{Im} \left\{ \int_0^1 \int_0^1 s_0(t) s_0(\tau) e^{-j\frac{\pi}{2}(t+\tau)} [\text{sinc } \pi B(t-\tau+1) + \text{sinc } \pi B(t-\tau-1)] dt d\tau \right\} \end{aligned} \quad (42)$$

where $\text{sinc } x \triangleq \sin x / x$. Furthermore, it is straightforward to show that

$$\int_{-\infty}^{\infty} G(f)df = 1 \quad (43)$$

and thus η is given directly by (42).

The maximization of (42) subject to the energy constraint of (41) has been carried out numerically using the MATLAB® optimization toolbox function “fminunc” (BFGS quasi-Newton method of convergence). In particular, for each value of B (BT_b if $T_b \neq 1$) the optimum complex signal $s_0(t)$ (represented by N uniformly spaced samples in the interval (0,1) is determined from which the fractional out-of-band power $1 - \eta$ is calculated using (42) for η . Because of complexity issues involved in computing the optimum solution, the number of sample points N is limited to 64. Furthermore, since the Gaussian integration required to evaluate with high accuracy the double integral of (42) requires a much higher density of sample values (not necessarily uniformly spaced), then to allow for Fourier interpolation we assume the signal to be bandlimited⁶ to the Nyquist rate, i.e., 32 ($32 / T_b$ if $T_b \neq 1$). Because of this bandlimiting assumption, certain optimum signal waveforms (particularly those at small values of B) which exhibit a sharp discontinuity will have a ringing behavior. This ringing behavior can be minimized by additional interpolation (filtering) but has proven difficult to eliminate completely.

Figs. 4a-f are plots of the optimum real and imaginary parts of $s_0(t)$ versus t at

⁶Of course, in reality the continuous time-limited signal $s_0(t)$ would have infinite bandwidth.

distinct values of B in the interval $0 \leq B \leq 3$. For small values of B we observe that the real part of $s_0(t)$ has sharp discontinuities at $t=0$ and $t=1$ and thus exhibits the ringing behavior alluded to above. As B increases the sharpness of the discontinuity at the edges diminishes and in the limit of large B , both the real and imaginary parts of $s_0(t)$ approach a sinusoidal behavior with unit period. Specifically, $s_0(t)$ tends toward the form $-\alpha_1 \sin 2\pi t + j(\beta_1 + \alpha_2 \cos 2\pi t)$ where $\alpha_1, \alpha_2, \beta_1$ are constants that also must satisfy the unit energy constraint of (40), i.e., $\beta_1^2 + \frac{1}{2}(\alpha_1^2 + \alpha_2^2) = 1$. Fig. 5 is the corresponding plot of optimum (minimum) fractional out-of-band power versus B . Also shown are corresponding results for MSK and SFSK modulations which can readily be found in [4, Fig. 2.11]. We observe that by optimizing the signal set at each value of B without loss in d_{\min}^2 or finite decoding delay performance, we are able to obtain a significant improvement in bandwidth efficiency. The quantitative amount of this improvement is given in Table 1 for the 99% and 99.9% bandwidths corresponding respectively to the -20 dB and -30 dB out-of-band power levels.

Table 1

Bandwidth Efficiency Performance of
TCM with Prescribed Decoding Delay

	Optimum Signals		MSK
	$\nu = 1$	$\nu = 2$	
$1 / B_{99} T_b$ (bps / Hz)	0.896	1.23	0.845
% Improvement over MSK	6.04	45.6	
$1 / B_{99.9} T_b$ (bps / Hz)	0.659		0.366
% Improvement over MSK	79.7		

4.2 Memory Two Case

Analogous to what was done for the memory one case, we need to maximize the fractional in-band power of (40) using now (31) for $G(f)$. Expressing the various Fourier transforms of (31) in terms of their associated signal waveforms and then performing the integration on frequency between $-B/2$ and $B/2$ as required in (40) produces the following result (again normalizing $T_b = 1$):

$$\int_{-B/2}^{B/2} G(f) df = \sum_{i=1}^6 P_i \quad (44)$$

where

$$\begin{aligned}
P_1 &= \frac{B}{4} \int_0^1 \int_0^1 (s_1^{(2)}(t) + s_2^{(2)}(t)) (s_1^{(2)}(\tau) + s_2^{(2)}(\tau))^* e^{-j\frac{\pi}{2}(t-\tau)} \text{sinc } \pi B(t-\tau) dt d\tau \\
P_2 &= \frac{B}{4} \int_0^1 \int_0^1 (s_0^{(2)}(t) - s_2^{(2)}(t)) (s_0^{(2)}(\tau) - s_2^{(2)}(\tau))^* e^{-j\frac{\pi}{2}(t-\tau)} \text{sinc } \pi B(t-\tau) dt d\tau \\
P_3 &= \frac{B}{4} \int_0^1 \int_0^1 (s_0^{(2)}(t) - s_1^{(2)}(t)) (s_0^{(2)}(\tau) - s_1^{(2)}(\tau))^* e^{-j\frac{\pi}{2}(t-\tau)} \text{sinc } \pi B(t-\tau) dt d\tau \\
P_4 &= 2 \text{Re} \left\{ \frac{B}{4} \int_0^1 \int_0^1 (s_0^{(2)}(t) - s_2^{(2)}(t)) (s_1^{(2)}(\tau) + s_2^{(2)}(\tau))^* e^{-j\frac{\pi}{2}(t-\tau+1)} \text{sinc } \pi B(t-\tau+1) dt d\tau \right\} \\
P_5 &= 2 \text{Re} \left\{ \frac{B}{4} \int_0^1 \int_0^1 (s_0^{(2)}(t) - s_1^{(2)}(t)) (s_0^{(2)}(\tau) - s_2^{(2)}(\tau))^* e^{-j\frac{\pi}{2}(t-\tau+1)} \text{sinc } \pi B(t-\tau+1) dt d\tau \right\} \\
P_6 &= 2 \text{Re} \left\{ \frac{B}{4} \int_0^1 \int_0^1 (s_0^{(2)}(t) - s_1^{(2)}(t)) (s_1^{(2)}(\tau) + s_2^{(2)}(\tau))^* e^{-j\frac{\pi}{2}(t-\tau+2)} \text{sinc } \pi B(t-\tau+2) dt d\tau \right\}
\end{aligned} \tag{45}$$

From the constraint in (37b), $s_2^{(2)}(t)$ can be expressed in terms of $s_0^{(2)*}(t)$ and then substituted in (45). Thus, the optimization problem reduces to finding only two signals, $s_0^{(2)}(t)$ and $s_1^{(2)}(t)$ by joint maximization of (44) combined with (45). (Note that $s_3^{(2)}(t)$ can be found from $s_3^{(2)}(t) = s_2^{(2)}(t) - s_0^{(2)}(t) + s_1^{(2)}(t)$ once $s_0^{(2)}(t)$ and $s_1^{(2)}(t)$ are determined).

Superimposed on Fig. 5 are the optimum fractional out-of-band power results for the memory two case. Due to the extremely time-consuming nature of the computer algorithms that perform the joint optimization procedure, particularly at low levels of fractional out-of-band power where extreme accuracy in satisfying the constraints is required, only results corresponding to values of $BT_b < 1$ (or equivalently $B < 1$ for $T_b = 1$) have been obtained thus far. Nevertheless, we are able to extract from these results the bandwidth efficiency improvement relative to MSK for the 99% (-20 dB) out-of-band power level and this is included in Table 1. We observe that there is a significant improvement in out-of-band power performance with no power efficiency penalty by going from a memory one (1 bit decoding delay) modulation to one that has memory two (2 bit decoding delay).

Acknowledgment

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

1. Q. Li and B. E. Rimoldi, "Bandwidth-efficient trellis-coded modulation scheme with prescribed decoding delay," *ISIT 1997*, Ulm, Germany, June 29–July 4, 1997. Full version to be submitted for publication.
2. G. D. Forney, Jr. "The Viterbi Algorithm," *Proc. IEEE*, vol. 61, no. 3, March 1973, pp. 268–278.
3. B. E. Rimoldi, "A decomposition approach to CPM," *IEEE Trans. Inform. Theory*, vol. IT-34, May 1988, pp. 260–270.
4. M. K. Simon, S. M. Hinedi and W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1995.
5. M. K. Simon, "A generalization of MSK-type signaling based upon input data symbol pulse shaping," *IEEE Trans. Comm.*, vol. COM-24, no. 8, August 1976, pp. 845–856.
6. F. Amoroso, "Pulse and spectrum manipulation in the MSK format," *IEEE Trans. Comm.*, vol. COM-24, no. 3, March 1976, pp. 381–384.
7. A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, New York, NY: McGraw-Hill, Inc., 1979.

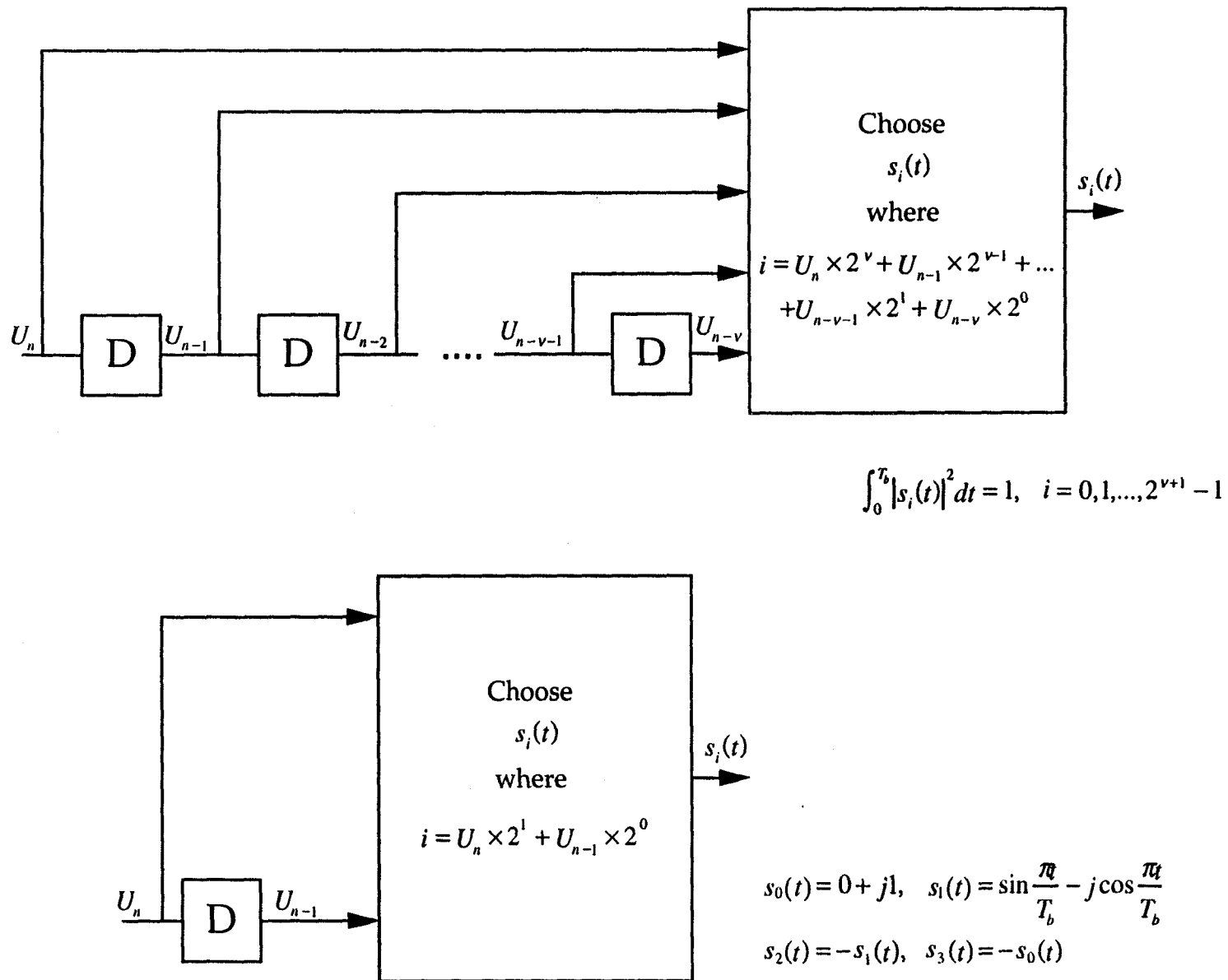
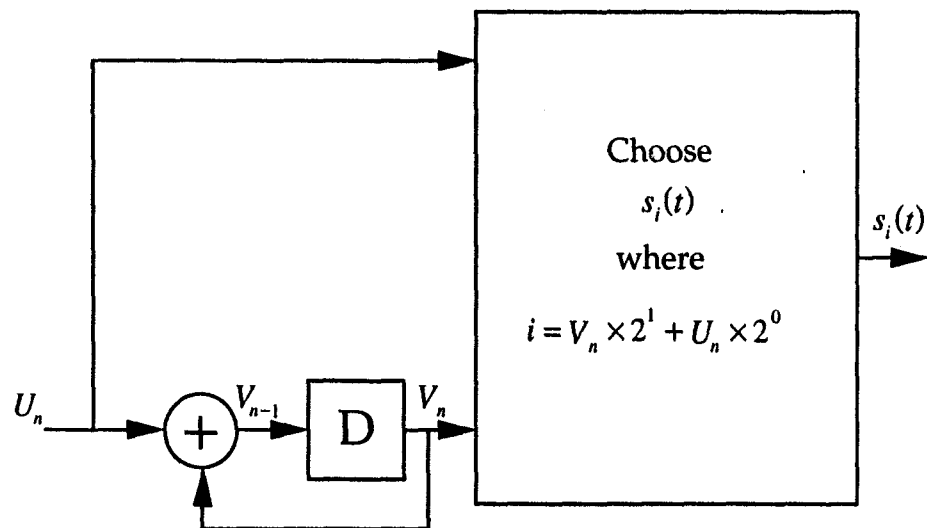


Fig. 1. (a) Trellis-coded modulation complex baseband transmitter (b) Special case of "MSK" ($v = 1$)



$$s_0(t) = 0 - j1, \quad s_1(t) = \sin \frac{\pi t}{T_b} - j \cos \frac{\pi t}{T_b}$$

$$s_2(t) = -s_0(t), \quad s_3(t) = -s_1(t)$$

Fig. 2. Trellis-coded modulation complex baseband transmitter for MSK based on Rimoldi decomposition of CPM

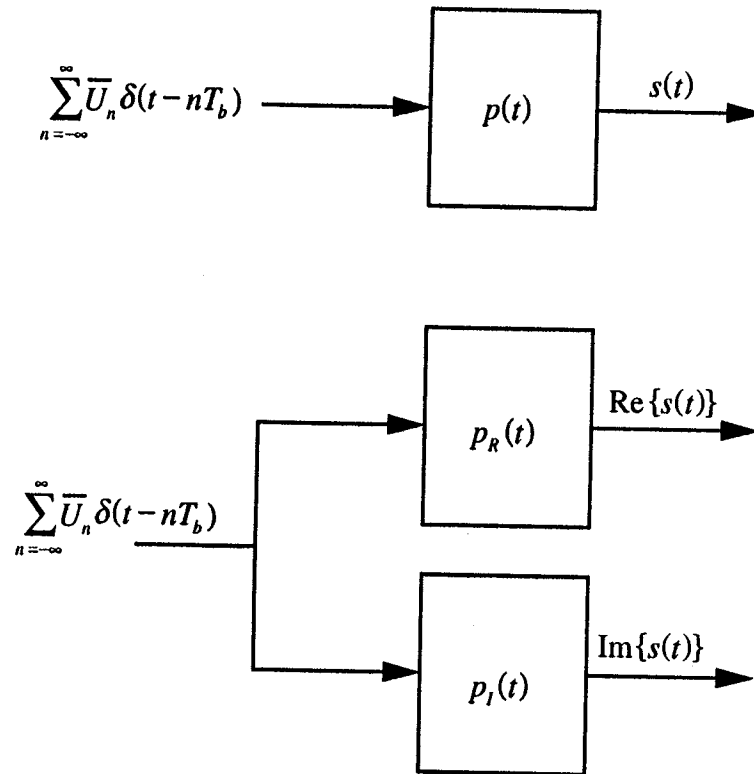


Fig. 3. (a) Complex baseband transmitter for MSK equivalent to Fig. 1 (b) I-Q baseband transmitter for MSK equivalent to Fig. 1

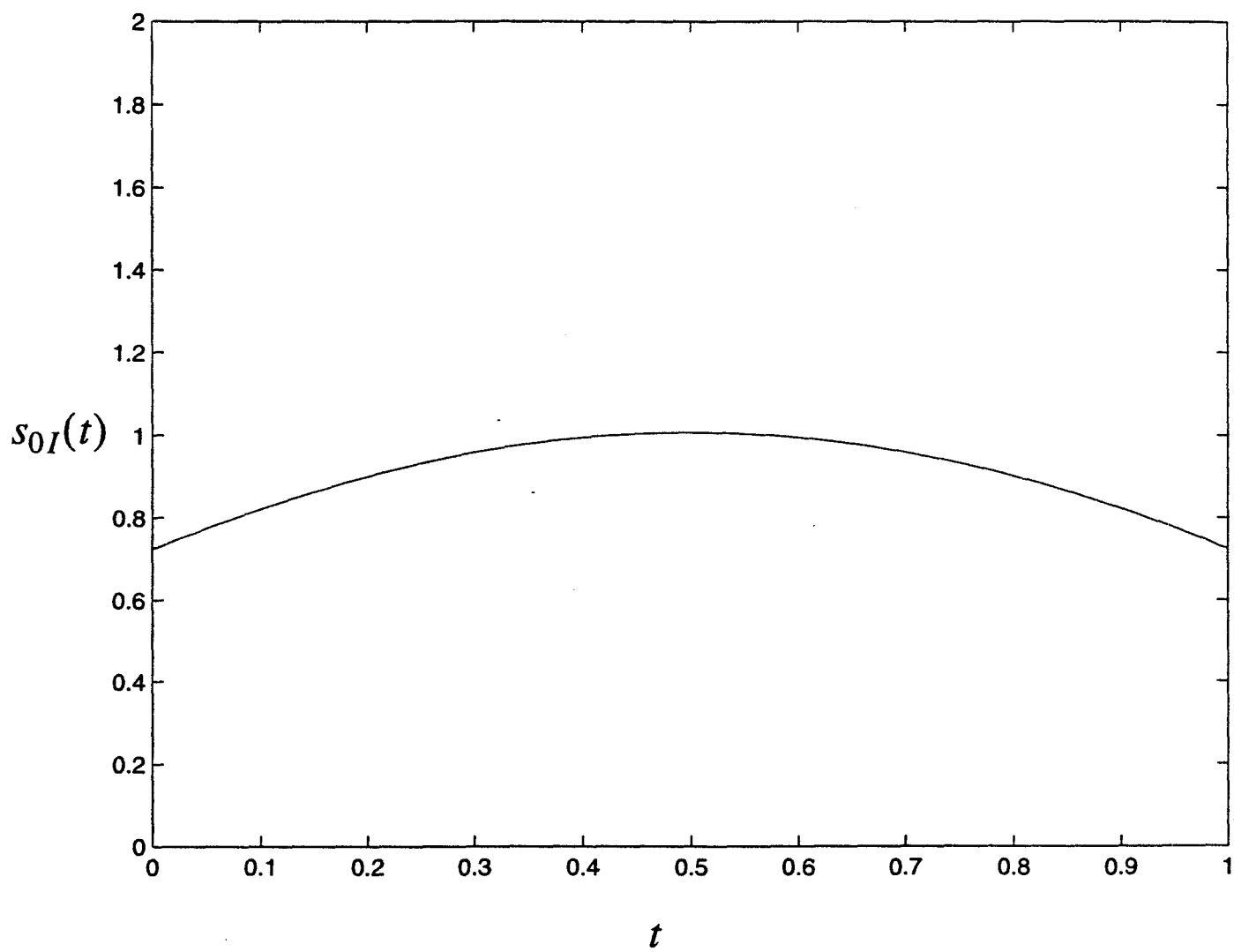


Fig. 4a Imaginary part of the optimum signal for bandwidth-time product = 0.2

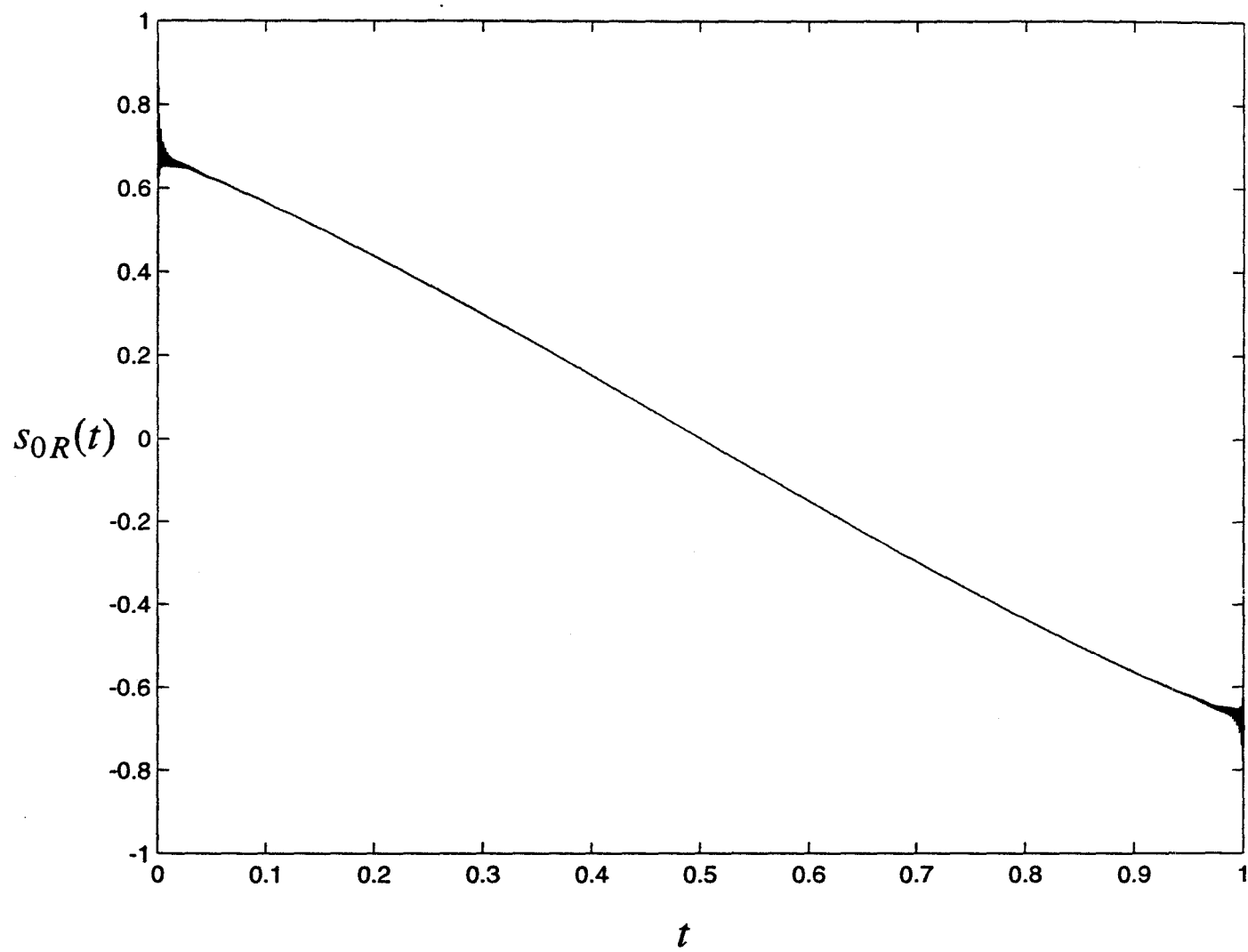


Fig. 4b Real part of the optimum signal for bandwidth-time product = 0.2

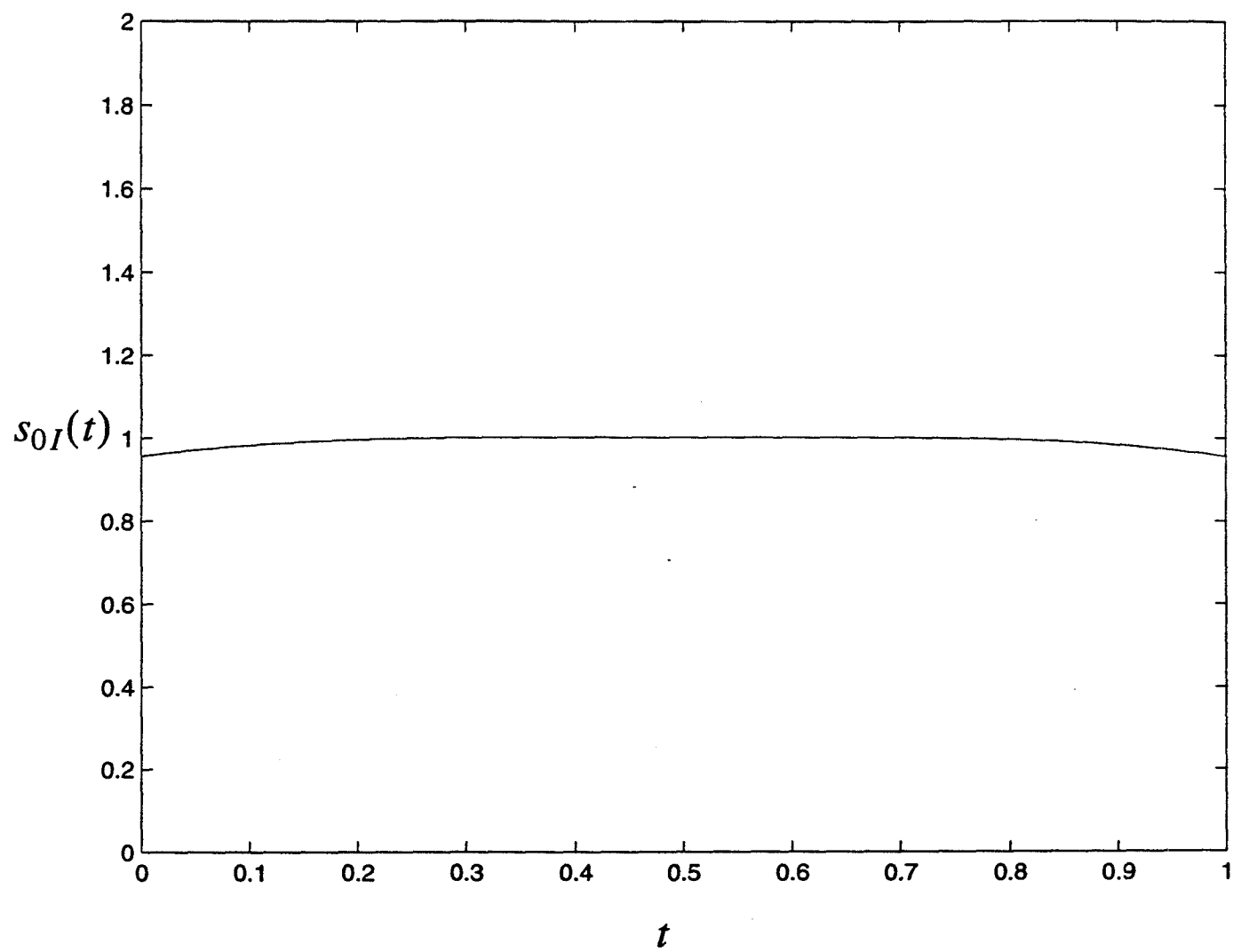


Fig. 4c Imaginary part of the optimum signal for bandwidth-time product = 1.0

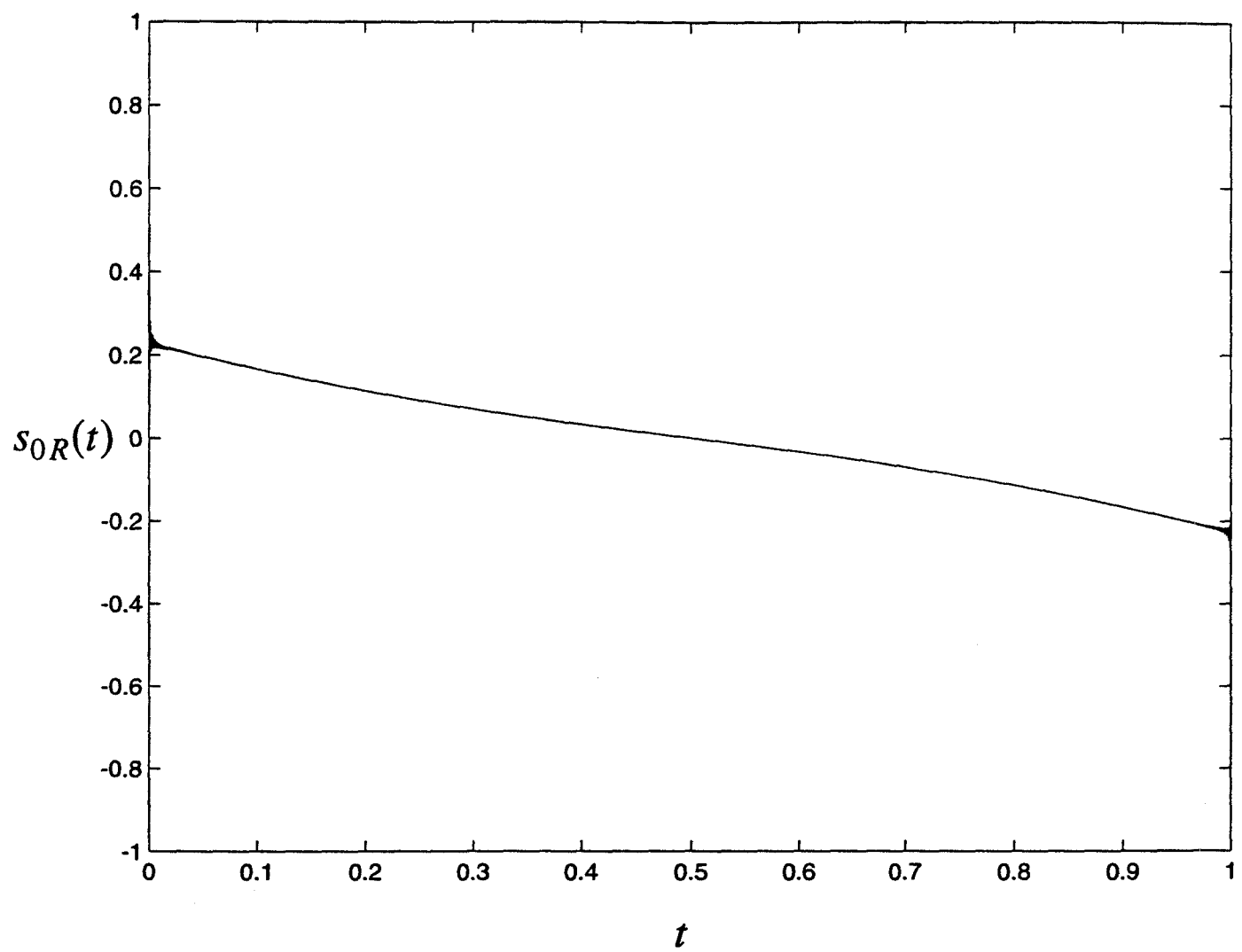


Fig. 4d Real part of the optimum signal for bandwidth-time product = 1.0

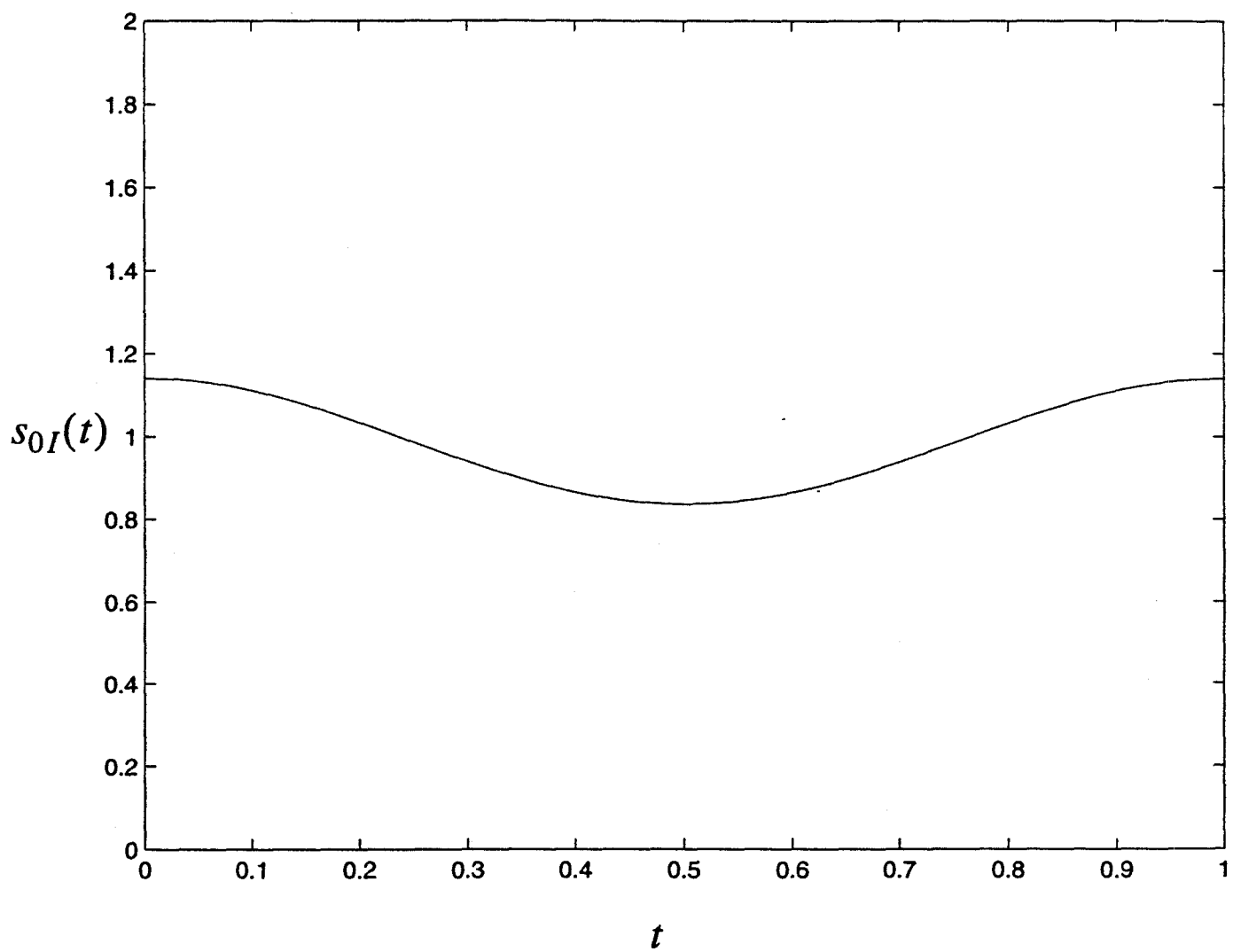


Fig. 4e Imaginary part of the optimum signal for bandwidth-time product = 1.8

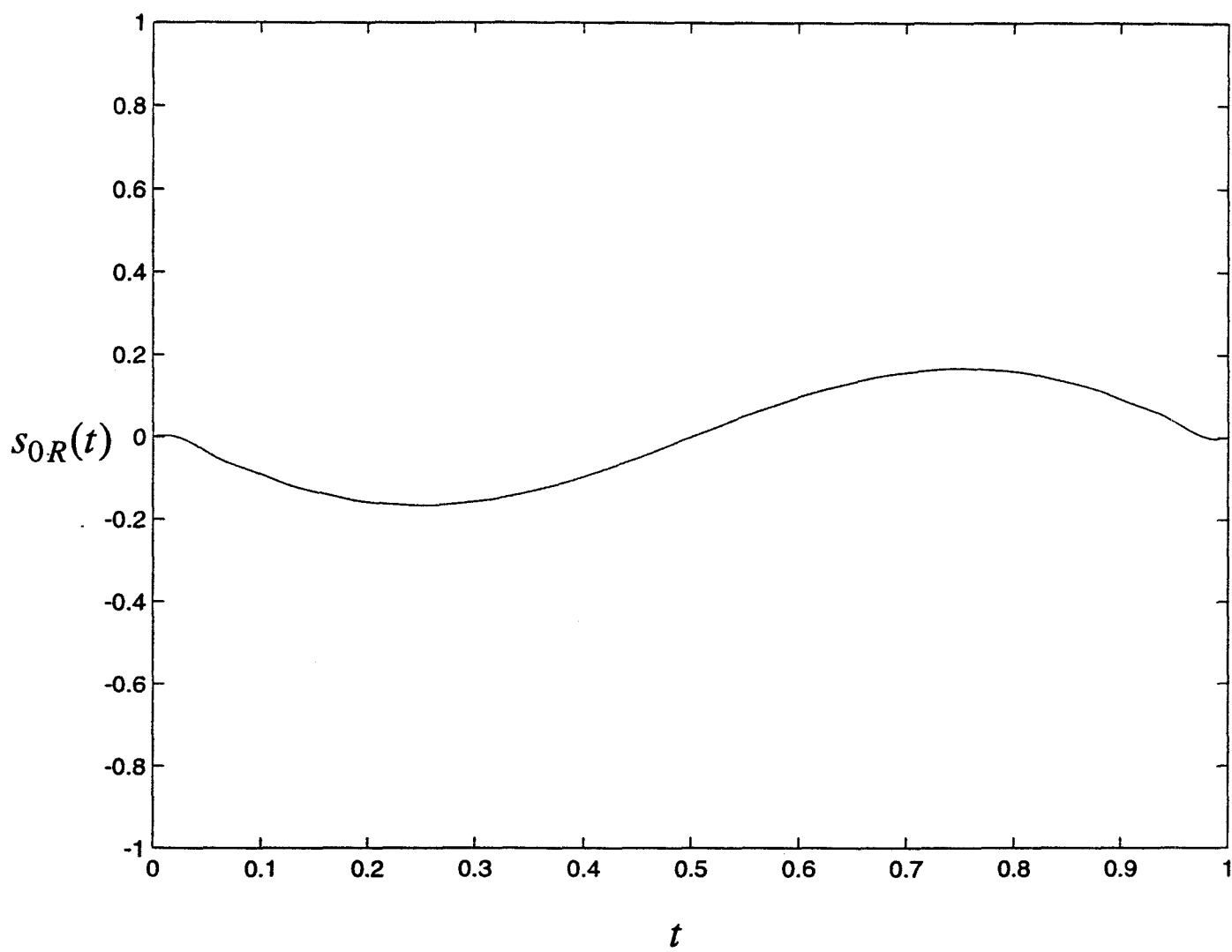


Fig. 4f Real part of the optimum signal for bandwidth-time product = 1.8

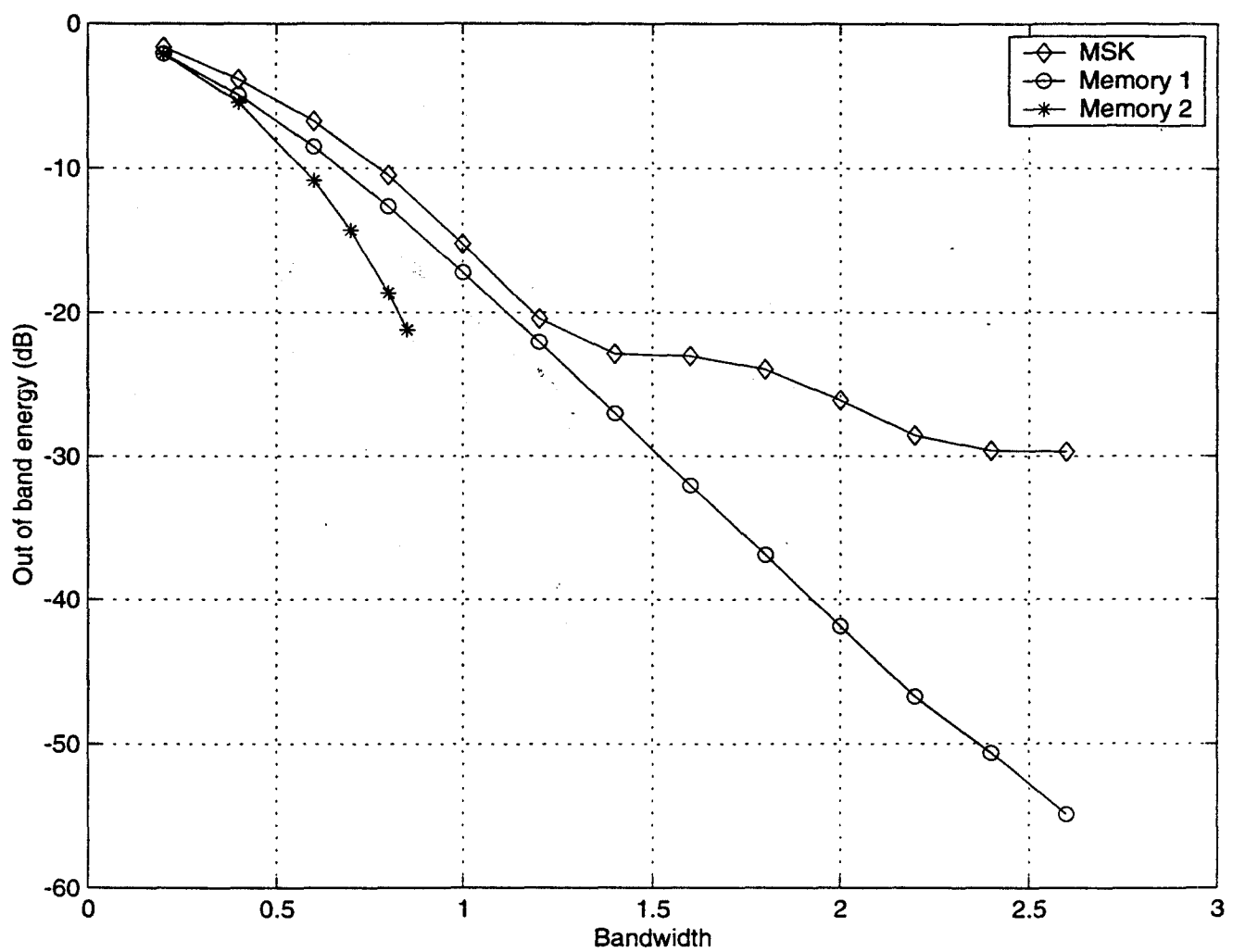


Fig. 5 Comparison of Fractional Out-of-Band Powers